Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 160 (2014) 205 – 214

Procedia
Social and Behavioral Sciences

XI Congreso de Ingenieria del Transporte (CIT 2014)

Urban bus demand forecast at stop level: Space Syntax and other built environment factors. Evidence from Madrid.

José Carpio-Pinedo[a] *

[a]*Dpto. de Urbanística y O. del Territorio, ETS Arquitectura, UPM. Avda. Juan de Herrera 4, 28040 Madrid, Spain.*

Abstract

Urban buses propose a challenge for traditional four-steps models of ridership estimation, as they require a different, closer scale approach, including the consideration of multiple possible stop-choices by travelers within walking distance. Thus, any model based on zoning and the bias of associating population to the nearest stop does not seem coherent in the case of urban bus.

This study empirically examines the potential of possible 'attraction' descriptors, such as spatial integration (as described by the Space Syntax methodology) and other urban environment factors in order to estimate urban buses ridership by a direct forecast model based on multiple linear regression. Common explanatory factors found in the literature include population and employment in the vicinity area, as well as transport system service and performance. Some authors have claimed the predictive power of built environment variables (summarized by Cervero and Kockelman's three Ds: Density, Diversity and Design), which are supposed to describe pedestrian accessibility and attractiveness. This paper proposes that spatial-configurational measures (e.g. Space Syntax) could play an important role, given that these factors have proved themselves synthetic proxies for many urban processes and in order to describe spatial-configurational hierarchy and consequent attraction power.

A demand forecast model at a stop level is explored by means of multiple linear regressions. Bus transport ridership at 84 stops in Madrid is forecasted using urban environment and spatial integration variables, as well as transport network accessibility indicators. Results seem encouraging and support that Space Syntax and other network integration variables could be an important asset for urban bus demand forecast models at a station level.

Peer-review under responsibility of CIT 2014.

* Corresponding author: Francisco J. Lamíquiz Daudén. Tel.: +34 91 336 65 92 ; fax: +34 91 336 65 34.
E-mail address: francisco.lamiquiz@upm.es

Peer-review under responsibility of CIT 2014.
doi:10.1016/j.sbspro.2014.12.132

Keywords: urban bus ; demand forecast at station level ; Space Syntax ; multiple linear regression ; Madrid

1. Introduction: urban bus and limitations for ridership forecasting.

Bus within dense urban environments is the transport mode that provides the most intensive level of coverage and gets the closest to every location in a city. This is crucial to ensure accessibility and the possibility of almost-door-to-door journeys for all citizens, in particular persons with reduced mobility. For this reason, urban buses play an important role for social goals in transport policy.

However, from a technical point of view, mobility by urban bus has represented a challenge for precise analysis and simulation at a closer level. The urban bus scale of service –usually less than 500 metres between two consecutive bus stops– does not allow ridership prediction at stop level by employing traditional transport models, such as four-step models (trip generation, distribution, mode choice and route assignment).

Being considered the most precise, reliable tools to predict ridership, these models are designed to describe and analyze mobility at a particular scale of zoning. The scale of metropolitan transport models usually enables to perform fine forecasts at a station level of a certain hierarchy, which relates to modes such as railways or underground (see figure 1). Errors are implicitly common at a closer scale.

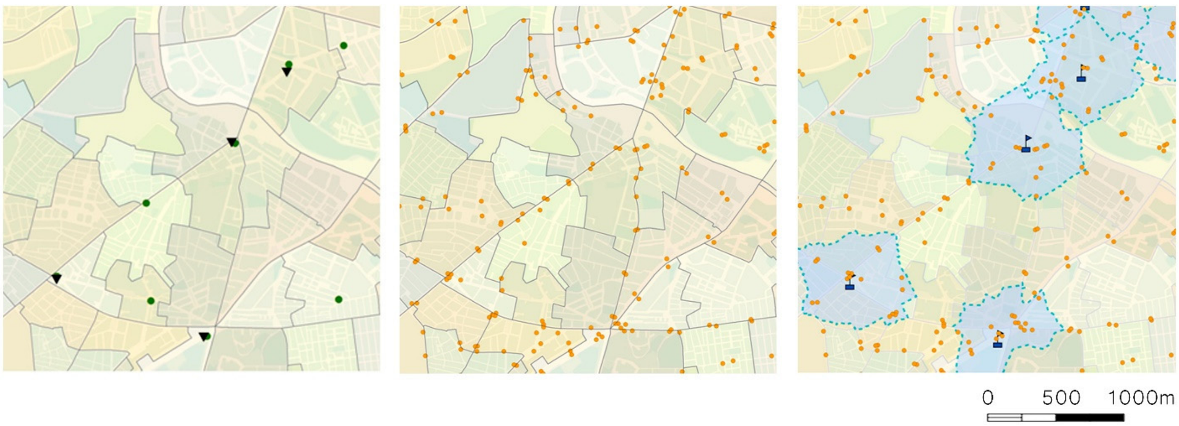

Figure 1: A detail of Madrid Transport Zones and railway/underground stations (left), bus stops (middle) and a 500metres pedestrian shed from some points - network distance (right).

In theory, in order to apply four-step models to the case of urban bus, the short distance between stops would need a very detailed scale of zoning, vast data and computing requirements, which may be unaffordable.

Moreover, the common belief that the more 'micro', the more detailed, may be misleading and contain conceptual and spatial incoherencies (Wegener, 2011). Short distances between stops allow pedestrians to reach multiple stops and choose. Thus, population cannot be directly assigned to the closest bus stop and any zone delimitation would not fit to real dynamics.

Direct models based on multiple regression analysis have tried to constitute a quick-response and less expensive alternative to the four-step demand forecast model: they propose a complementary approach to estimate ridership as a function of transit services features and station urban environment (Gutierrez et al., 2011), summarized by Cervero and Kockelman's three Ds (Density, Diversity and Design), which are more simple, available data. However, both approaches are usually based on zoning or have an implicit bias of direct assignment of population to the closest stop, as this is indeed a logical assumption (and has a good performance) in other modes such as underground. Nonetheless, for the case of urban buses, the availability of multiple stops at a pedestrian shed urges to consider the preferred choice of bus-riders, according to differences of 'attraction' between stops. This paper explores this question by analyzing the potential of spatial integration measures –Space Syntax methodology–, transport network accessibility and urban environment features in order to constitute a direct forecast model based on linear multiple regression.

In section 2, the goals of this study are settled. In section 3, some exploratory hypotheses are set, in order to constitute a conceptual basis for a methodology, which is described in section 4. In section 5, still exploratory statistical results are presented and analyzed. Finally, section 6 drafts provisional conclusions and discusses a research agenda to go further in the future.

2. Goals of this study.

The main goal of this study is to empirically explore what urban variables are more associated to bus ridership, with a special focus on spatial integration measures and other possible 'attraction' indicators. To achieve this goal, bi-variable linear regression analysis are the most appropriate method.

The ambition lying under this goal is to build a direct model to forecast bus ridership at stop level, based on multiple regression analysis. This is adjusted for the case of Madrid city.

3. Hypotheses and conceptual framework: factors related to bus stop attraction.

In cases where pedestrians can reach two or more bus stops at a distance that they are willing to walk, what factors would be decisive for their final choice?

Some exploratory hypotheses are briefly proposed below, based on diverse literature and common wisdom and experience as transport users.

A. First of all, minimum metric distance to the bus stop may be crucial for most travellers, no matter what other factors might affect. → stop #1 in Figure 2.
B. Secondly, configurational studies initiated by Space Syntax theory and methodology (Hillier&Hanson,1984; Hillier, 1996) claim that pedestrians do not seek to reduce the physical distance, but the topological-cognitive one. Literature on Space Syntax is impressive in quantity and in results: spatial configuration, based on topological distances, has proved to explain the location and distribution of many urban dynamics but, most especially, pedestrian flows, walkability and pedestrian routes choice.
   Pedestrians may walk to the "nearest" stop in topological terms (minimum number of "turns", not "metres"), which relates to mental effort instead of physical effort. → stop #2 in Figure 2 implies walking more physical distance, but less direction change.
C. Moreover, since riding a bus usually involves waiting first at an exposed public space,

it may be decisive–consciously or unconsciously – to wait in locations where one feels comfortable. Facilities and design features may play their role, but in this paper the focus is on comfort associated with urban planning. Comfort involves a feeling of both safety and entertainment. What urban variables are related to this in public spaces? Literature from classics (Jacobs, 1961) has basically associated this to the presence of other people, which has, moreover, a multiplier effect (Gehl, 1987). → stop #3 in Figure 2.

D. Finally, citizens may be willing to walk a longer distance if they get to a "more useful" node in terms of transport offer (more and better connections). → stop #4 in Figure 2.

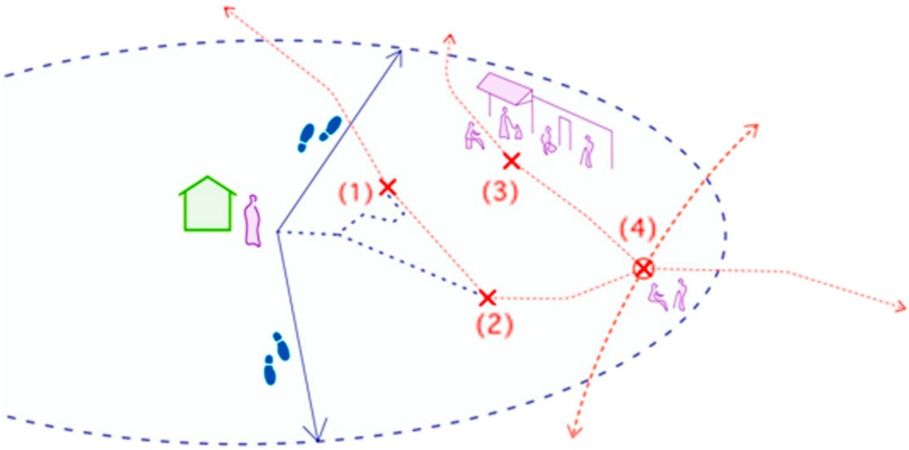

Figure 2: Multiple bus stops at a walking distance. Hypotheses of preferred choice.

The hypothesis that bus demand is associated with spatial integration (Space Syntax variables) is based on the following remarks:

- Space Syntax does not only relate to reason B, but also to reason C, given that the spatial hierarchy defines the distribution of pedestrian flows and, moreover, this is associated to the concentration of commerce and other services (Hillier, 1999).
- Also, in most cases, transport planners may have located stops and stations on major axes of the street network, as it actually happens in the case of Madrid. Moreover, this intuitive localisation of transport points is likely to be repeated at different scales: stations of metropolitan relevance on the most important axes of a city, whereas common bus stops on the major axe of a neighbourhood. So Space Syntax could be associated to reason D as well.

These hypotheses encourage to study and value the relevance of spatial integration variables as a proxy of many other variables (reason B+C+D) and analyze its explanation power for bus ridership forecasting.

In next section, a first exploratory methodology is proposed in order to evaluate these hypotheses.

4. Methodology.

4.1. Case study

The case study is the city of Madrid, the largest city of Spain. The population of the city is roughly 3.3 million and the entire population of the Madrid metropolitan area is calculated to be around 6.5 million.

The great development of urban bus network, combined with other transit modes (regional railways, regional buses, 13 lines of underground, etc.) and a wide diversity of urban environments make Madrid an ideal case for this study. According to official data, urban bus services include 217 routes, covering 3,901 km and 10,498 stops. Each route typically covers about 9 km, with 25 stops at intervals of 330 meters. The average journey distance per passenger is 3 kilometres.

4.2. Delimitation of urban environments.

During 2011 and 2012, the Madrid Transport Authority (*Consorcio Regional de Transportes de Madrid*, CRTM) elaborated a rich analysis of 128 transport nodes and their urban environment in order to support the City Council on the new Madrid Masterplan. Taking advantage of this work, 84 nodes with a sufficient amount of data are selected, once excluded the most important nodes of the study (airports, long-distance railway station, large interchanges, etc.).

From the transport network point, urban environment is defined as a pedestrian shed of 500 metres radius through road-network distances, which are more accurate than coverage circles based on Euclidean distance.

Radius of 500 metres is considered in the literature as a convenient pedestrian shed to capture proximity dynamics (Mashhoodi & Berghauser Pont, 2011). However, this all/nothing consideration of urban territory is not ideal and requires a finer consideration –through gravity equations or distance decay functions– in future versions of this work.

4.3. List of variables and sources of data.

Two twin-processes have been executed in parallel. These correspond to two target dependent variables to be explained, modelled and forecasted.

- One is the total urban bus demand at a particular transport node. Most of stops in urban areas give access to more than one line, and/or there is another stop round the corner or a few metres away. Since all these stops are within the same urban environment, a demand forecast model based on its characteristics would provide the same results. Thus, a first approach is to understand all these stops as a single system, so the variable to predict is “Total bus demand”.
- Also, accepted the approach above, it is logical that the number of lines will play a clear role in total demand, no matter the characteristics of urban environment. That is why a second variable to study is “Bus demand/number of lines”. This is a necessary approach, so that the influence of the number of lines on total demand could be isolated. However, it is no doubt rough to divide just as if all lines could be considered equally relevant. That is why at the end of this paper a pair of alternatives for future research are pointed.

The source of both variables is a demand report carried by TARYET for CRTM (2008), during a whole working day (7am-10pm). Variables correspond to the total daily demand (both on and off) and are the sum of all bus stops that can be perceived as a single public transport point: adjacent, across the street or round the corner.

A list of independent variables (table 1) is selected in order to reflect the hypotheses explored in Section 3.

A concentration of potential trip origins and destinations is related to reasons A and C. The number of residents around a station is usually considered the main factor of trip generation; whereas the number of workers is the main factor for trip attraction. Also, retail shops and other services constitute potential destinations at the area,but also contribute to a vibrant urban space. Moreover, if transport planning has followed a supply-demand rule, this variables may be directly related to reason D as well.

- number of inhabitants.
- number of jobs,
- number of retail shops and other services.

Table 1: List of variables included.

Group of variables	Variables	Source of data
Urban environment: Population Employment	Population at 1000m (Netw.dist.) Population at 300m (Netw.dist.) Population at 500m (Netw.dist.) Jobs at 500m (Netw.dist.)	Madrid Region Statistical Institute. (IESTADIS) Data provided to CRTM.
Population reach	Population reach (30m travel) Population reach (60m travel)	Madrid Statit. Institute (IESTADIS) + Network analysis by CRTM.
Urban environment: Retail shops and services. *(classification is explained in Carpio-Pinedo, 2014)*	Retail shops and services: Total (500m ND) Basic need retail shops (500m ND) Basic need retail shops _ new formats (500m ND) Basic need services (500m ND) Occasional retail shops (500m ND) Occasional services (500m ND) Bars, restaurants and others (500m ND)	Madrid Region Statistical Institute. Economic Activity Units Directory. (DUAE)
Urban environment: Spatial Integration (Space Syntax methodology)	Log _ Integration R3 (Ave) Log _ Integration R5 (Ave) Log _ Integration RN (Ave) Log _ Integration R3 (Max) Log _ Integration R5 (Max) Log _ Integration RN (Max)	Madrid Axial Model from National R&D Project 'La Ciudad Paseable' (The Walkable City). Lamíquiz y Pozueta, 2008.
Transport network: Number of lines.	# of lines: regional railways # of lines: underground # of lines: regional bus # of lines: urban bus	CRTM transport model.
Transport network: Network accessibility. Based on Curtis, 2011.	Metropolitan Betweenness City Closeness Metropolitan Closeness City Straightness Metropolitan Straightness	CRTM transport model.
Transport demand.	Total Demand (all modes) Regional Railways Demand Underground Demand Underground+bus Demand	CRTM data.

Also, passers-by contribute to liveability and comfort while waiting (reason C). As explained, pedestrian flows have been strongly associated to spatial configuration (described by Space Syntax measures), so it is closely related to reason B as well. Moreover, spatial configuration is a major explanation factor of retail location. Taking into account Space Syntax literature, spatial configuration variables considered for this study have been:

- Global integration – weighted mean (log) in the urban environment-radius 500m.
- Global integration – maximum value (log) in the urban environment-radius 500m.
- Local integration – weighted mean (log) in the urban environment-radius 500m.
- Local integration – maximum value (log) in the urban environment-radius 500m.

Finally, the level of utility of each particular stop within the transport system (reason D) can be described in different approaches:

- Number of lines of each mode: more lines imply a wider territory directly reachable.
- Public transport accessibility, based on Curtis (2011):
  - average travel times to reach all nodes in Madrid city → City Closeness.
  - average travel times to reach all nodes in Madrid region → Region Closeness.
  - average necessary transfers to reach all nodes in Madrid city → City Straightness.
  - average necessary transfers to reach all nodes in Madrid region → Region Straightness.
- Population reached in a 30 min journey.
- Population reached in a 60 min journey.
- Demand (on other modes). This is a direct indicator of utility, but also supposes potential bus demand after transferring.

4.4. Statistical methods.

In this first exploratory work, statistical methods are the most appropriate tools in order to manage and analyze this great amount of data.

First of all, bivariate correlation helps identify association between two variables. All variables are compared with total bus demand and bus demand/number of lines. In second place, a first draft bus demand forecast model at stop level is adjusted. This is done by multiple linear regression analysis. In this, demand variables are excluded, given that many cities do not have this kind of data and, as a result, the model would be less useful. On the other hand, there might be a possible risk of incoherence if the demand forecast model relies on other demand variables.

In a first attempt, the focus is on models that need population and employment data, as well as transport network variables: these are usually the most available or easy to calculate group of variables (**1.**). In a second approach, measures of retail and services concentration are included. This group of variables is not always available, so that it is convenient to value its relevance independently (**2.**).

Both approaches are repeated **A)** including spatial integration measures (Space Syntax) and **B)** excluding these variables. In this way, it is possible to measure the potential and weight of spatial integration factors in order to build a forecast model.

5. Results.

5.1. Most associated variables. Bivariate correlations.

The highest R correlation coefficients are presented in table 2. Some general comments can be done:

- Correlation coefficients R for total bus demand are higher than those for bus demand per line. This is an expected, logical result, given that dividing by the number of lines is a tough method to isolate the multiplying effect of the number of lines.
- Demand variables (other public transport modes) are the most associated factors with urban bus demand. This may point at a very close relationship between similar variables: being endogenous relationships, they may be co-influenced or inter-related; but also they may be the result of the same dependent exogenous factors.
- It is interesting that many urban features seem to be associated with urban bus demand, equally or even more intensely than transport network variables (except for the already discussed demand on other modes and the logical high correlation between total demand and number of lines).
    - The relevance of population is recurrent, although it is important to note that it is not only population in the vicinity area that counts (physical proximity), but also population reached in a given travel time by public transport.
    - Also, spatial integration variables (Space Syntax) prove their explanation power (R=0,519) on its own right. Two of the tested variables are consistently high-correlated with bus demand. Average global integration (accessibility within the whole urban system) and maximum local integration (highest value within the bus stop vicinity). These are expected results, since global integration describes metropolitan centrality vs. periphery aspects, which is considered by the research community as a key-factor for public transport ridership. Maximum local integration corresponds to the most accessible axe in the area, usually within the main, most connected street, which articulates pedestrian natural movements within the area.
    - Finally, it is very interesting to realize the rather high correlations between urban bus demand and the presence of certain kinds of retail and services. This may relate to the reasons already argued in section 3 and 4.3: these activities are potential trip destinations but also create attractive, liveable urban places.

Table 2: Best results of bivariate correlations.

Correlation with Total Bus Demand		R	**Correlation with Bus demand/#lines**		R
Tr. Demand	Reg.Railways Demand	0,916	*Tr. Demand*	Reg.Railways Demand	0,652
Tr. Demand	Underground Demand	0,775	*Tr. Demand*	Underground Demand	0,577
Tr.N: Lines	# of lines: urban bus	0,735	*UE: Population*	Population at 1000m (Netw.dist.)	0,384
Pop. Reach	Population reach (60m travel)	0,601	*Tr.Net. Accessibility*	City Closeness	0,382
Pop. Reach	Population reach (30m travel)	0,593	*Pop. Reach*	Population reach (30m travel)	0,357
UE: Retail & Serv.	Basic need services (500m ND)	0,587	*UE: Spatial Integr.*	Log _ Integration RN (Ave)	0,344
UE: Retail & Serv.	Occasional services (500m ND)	0,578	*Pop. Reach*	Population reach (60m travel)	0,343
UE: Spatial Integr.	Log _ Integration RN (Ave)	0,519	*UE: Population*	Population at 500m (Netw.dist.)	0,339
UE: Spatial Integr.	Log _ Integration R3 (Max)	0,518	*UE: Spatial Integr.*	Log _ Integration R3 (Max)	0,325
UE: Retail & Serv.	Basic need retail shops (500m ND)	0,476	*Tr. Demand*	Bus+Underground Demand	0,319
UE: Population	Population at 300m (Netw.dist.)	0,466	*Tr.Net. Accessibility*	Metropolitan Closeness	0,318
Tr.Net. Accessibility	Metropolitan Closeness	0,465	*UE: Spatial Integr.*	Log _ Integration RN (Max)	0,307
UE: Retail & Serv.	Retail shops and serv.: Total (500mND)	0,460	*UE: Retail & Serv.*	Basic need retail shops (500m ND)	0,307

5.2. Bus ridership forecast model. Multiple linear regression.

Table 3 summarizes the results of the four multiple linear regressions carried out (A1, A2, B1, B2) throught their coefficients of determinations (R2 and R2 corrected values), in order to check the achieved level of prediction.

Once again, the model for total bus demand performs better forecasts (R2 between 0,737-0,771) than the one for demand/line (0,521-0,574).

It is been confirmed that Space Syntax variables actually improve R2 coefficients, both in the case of total demand and demand/line (A1>B1; A2>B2). The inclusion of retail and services variables also improves R2 coefficients (A1>A2; B1>B2). However, the opposite can be said when checking the R2 adjusted values. This points out at an important issue, which is the representativeness of the sample, which might be biased by the CRTM selection. This selection of cases (84) seems very diverse, but it is clear that all cases correspond to “points of interest” for the Transports Authority –current or future strategic importance–.

Table 3 : Multiple linear regression. Coefficients of determination R2 summary.

	Inhabitants Jobs	Population reach	Retail and services	**Spatial integration**	Transport Network	Demand	Total Bus Demand R2	Total Bus Demand R2 adjusted	Bus demand (total /line) R2	Bus demand (total /line) R2 adjusted
A1							**,771**	**,634**	**,574**	**,321**
A2							**,746**	**,642**	**,558**	**,378**
B1							**,765**	**,664**	**,547**	**,351**
B2							**,737**	**,664**	**,531**	**,401**

6. Conclusions and research agenda.

Transport planners have not customarily attempted to forecast urban bus ridership at stop level, once this has been considered unaffordable based on common use methodology.

This first exploratory study proves that, despite the micro-scale that they seem to involve, direct models to forecast urban bus demand at stop level are achievable with rather realistic resources.

This could be a first step to build up a methodology for a better planning of urban bus itineraries and stop locations.

An underlying conclusion is that no vast work is necessary to get to micro if it is not coherent with the issue being faced, as Wegener (2011) claims. When access points to transport are distributed so densely that citizens can choose between two or more stops, it is clear that methodologies must be re-thought, especially if they are based on zonal divisions and the bias of assigning population to the closest stop.

The results of this study are hopeful and prove that new kinds of factors play their role at a closer urban scale. Based on a set of limited hypotheses, the list of variables proposed has helped to identify what urban variables are more associated to bus ridership.

Among these, it has been proved that spatial integration measures –as described by Space Syntax methodology– can become an important asset to predict bus demand.

Although results are encouraging, they require further development. Some of the limitations that must be overtaken in future versions of this work are:

- A sample that is not biased by transport planners and may reduce the gap between R2 and R2 adjusted values.
- A better consideration of population, employment, retail and services in the vicinity area of the bus stop, by taking advantage of gravity functions or "distance-decay" formulas (Gutiérrez et al., 2011).
- A more precise way of isolating the number of lines, without a direct division. This could be done by using disaggregated data of each line and by evaluating its role within the transport system independently. A possible way is to describe its utility by the population directly covered by each line. Another possibility is to pre-estimate the demand of the line as a whole with four-step models and then, define the distribution of this demand at stop level with a model based on urban environment factors. The later would make the whole method depend on traditional models once again and, thus, it would be less useful independently.
- Socio-economic data of population could enrich the model, since all studies claim that urban buses are the preferred transport mode for a particular group of population, such as senior citizens. However, there are two recurrent issues to deal with: a) the availability of socio-economic information, and b) if available, the aggregation scale of this data. Would it be coherent with the scale of urban bus stops?

Acknowledgements

I would like to thank the valuable ideas and comments that I have received from an anonymous reviewer and from several colleagues from Consorcio Regional de Transportes de Madrid: Antonio García Pastor, Domingo Martín, Francisco Herranz, Maite Antón and Francisco Gómez.

References

Carpio-Pinedo, J. (2014) "Localización y evolución del comercio y servicios a pie de calle en los entornos urbanos: factores locales frente a la multi-accesibilidad. El caso de Madrid." Territorios en Formación, nº6.

Curtis, C. (2011). "Integrating Land Use with Public Transport: The use of a discursive accessibility tool to inform metropolitan spatial planning in Perth". Transport Reviews, vol.31, no. 2.

Gehl, J. (1987). Life between buildings: using public space. Nueva York, Van Nostrand Reinhold.

Gutiérrez, J., Cardozo, O. D., & García-Palomares, J. C. (2011). Transit ridership forecasting at station level: an approach based on distance-decay weighted regression. Journal of Transport Geography, 19(6), 1081-1092.

Hillier, B., & Hanson, J. (1984). The social logic of space. Cambridge: Cambridge University Press.

Hiller, B. (1996). Space is the Machine: A configurational theory of architecture. Cambridge: Cambridge University Press.

Hillier, B. (1999). Centrality as a process: accounting for attraction inequalities in deformed grids. Urban Design International, 4(3-4), 107-127.

Jacobs, J. (1961): The death and life of great American cities. Nueva York: Random House.

Lamíquiz, F.J. y Pozueta, J. (2008) Informe Científico: "Análisis de la influencia de los rasgos urbanísticos en la movilidad peatonal". Plan Nacional I+D+i, subprograma PEIT, CEDEX, Ministerio de Fomento. Documento de trabajo sin publicar, DUyOT, ETSAM, UPM

Mashhoodi, B., & BerghauserPont M. (2011) "Studying land-use distribution and mixed- use patterns in relation to density, accessibility and urban form." ISUF, Montréal, Canada, August.

TARYET for CRTM (2008): "Estudio de determinación y caracterización de a demanda en día laborable en las líneas diurbans de la red de EMT de Madrid (2007)". Informe no público. Consorcio Regional de Transportes de Madrid.

Wegener, M. (2011). From Macro to Micro—How Much Micro is too Much?.Transport Reviews, 31(2), 161-177.